Applying Features Based on Word Embedding Techniques to 1D CNN for Natural Disaster Messages Classification

Mohammad Reza Faisal, Irwan Budiman, Friska Abadi, Dodon Turianto Nugrahadi, Muhammad Haekal
Department of Computer Science
Lambung Mangkurat University
Banjarbaru, Indonesia
reza.faisal, irwan.budiman, friska.abadi, dodonturianto@ulm.ac.id, muhammadhaekal77@gmail.com

Indrajani Sutedja
Information Systems Department
Bina Nusantara University
Jakarta, Indonesia
indrajani@binus.ac.id

Abstract:

Messages of natural disasters on social media can be used as an early warning and mitigation of natural disasters. Researchers developed a machine learning-based classification model to identify natural disaster messages automatically. In comparison, the Previous research used shallow learning classification algorithms and feature extraction techniques with vector space representation. This feature extraction technique produces high-dimensional data. This technique eliminates word order information also so that it will lose the meaning of the sentence. Word embedding is a method for transforming the word into vectors with numeric values that capture a word's semantic and syntactic information. We use this method to generate structured data that keep word order, semantic, and syntactic information. The generated data are processed using deep learning, which is 1D CNN. This learning method is generally applied in signal classification, so we must study how to determine the input for 1D CNN to get the best accuracy. We use several techniques to resize the number of words and three-word embedding techniques, i.e. word2vec, Glove and fastText. We find that mean and word2vec are the resized number of word and word embedding techniques that can give the best accuracy to classify natural disaster messages.

Keywords— social media, natural disaster, word embedding, CNN, classification